Designing From Data: A Case Study From the Health Mashups Service

Frank Bentley
Yahoo! Labs
Sunnyvale, CA USA
fbentley@yahoo.inc.com

Konrad Tollmar
Wireless@KTH, Royal Institute of Technology (KTH)
Kista, Sweden
konrad@kth.se

Abstract
Instrumentation of mobile service interactions now allows for the ability to understand use in new ways and to base the design of system iterations as well as future services on data obtained from use "in the wild" over extended periods of time. This position paper will detail a few specific instances in the Health Mashups project where we were able to use usage data to improve the design and to create design implications to increase engagement in new services.

Author Keywords
Health; Instrumentation; Design

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

MobileHCI 2013, Aug 27 – 30, 2013, Munich, Germany.
ACM XXX-X-XXXX-XXXX-X/XX/XX.

Introduction
The instrumentation of user interactions is quickly changing a designer's ability to understand and react to the use of a deployed system in the wild. While much previous work in the Mobile HCI and Ubicomp fields was conducted at a small scale (often 10-12 users) with minimal usage data (e.g. [1,4]), new larger deployments are allowing for the interpretation of statistically significant usage data in near real time and in much tighter iteration cycles.

While much has been written about using this type of data in A/B testing to pick better designs based on quantifiable outcomes (e.g. more time on site, higher ad click-throughs, etc.), comparatively less has been written on combining large-scale usage data with qualitative data to create new inspirations for design and to design and evaluate iterations for systems. This position paper will describe how we applied this process at the Motorola Mobility Applied Research Center and Wireless@KTH to iterate on the Health Mashups service, to verify that our design changes achieved the desired effect, and to create inspiration for the design of new mobile services.

The Health Mashups Service
Health Mashups is a mobile service intended to help users understand the complex interactions between

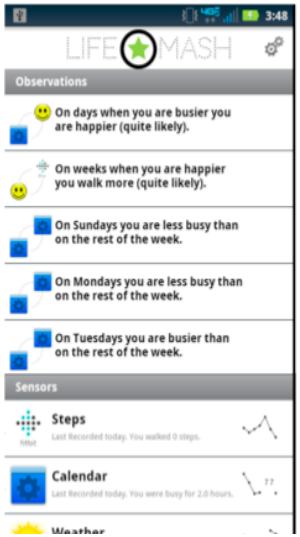

Figure 1: The main observations screen in the Health Mashups application with examples of the types of observations that were shown.

various aspects of their wellbeing and their context. It collects a variety of data from sensors such as the FitBit pedometer/sleep sensor and WiThings WiFi scale. In addition, it includes automatically-logged context from the phone such as location, weather for that location, and time busy in a user's calendar as well as manually logged pain, mood, and food intake information. The system analyzes trends and interactions between these data streams over long time periods and presents statistically significant results to users in the form of a feed of natural language sentences. For example, the feed (see Figure 1) can contain observations such as "On Wednesdays you walk significantly less than on other days." or "On days when you are busy, you eat significantly more."

A pilot study was performed with 10 users for 2 months and is reported in [2]. Both qualitative and quantitative data from this study were used in a design iteration as discussed below leading to a larger trial with 60 users over 90 days and an additional 52 users recruited online who were given access to the system but were not a part of the interviews in the trial.

Using Data for Design Iterations

Several aspects of our system design were able to benefit from instrumentation in both the smaller pilot study and larger deployments that led to design changes.

In our pilot study, we saw quite low usage of the manually logged data streams of food and exercise. We also saw a generally low use of the application itself in reviewing the significant observations that were created. These insights came from a combination of our interviews, where participants talked about forgetting to log, placing the widget on a secondary home screen, and not knowing when new observations were visible. This, combined with the quite low usage of manual logging (see Figure 2), gave us ideas on what would need to modified to increase engagement.

Based on these observations, we made several design changes to increase engagement, and through increased logging, increased the quality of the observations that the service could generate. First, we added a reminder system [3] for logging. When the user first entered the application, they would be shown a setup screen with each data stream. Manually logged streams, such as food and mood, involved setting a time when an Android notification would appear in the status bar each day with one click access to log the item of interest.

We also added Android notifications that would appear whenever a new statistically significant observation was generated for a user. Thus, they did not have to scroll through a long list wondering what was new, but rather they could let the system be proactive in telling them when they should engage.

Finally, once the larger trial was underway, we began to observe the quantity of different types of observations that were generated for users. There were often multiple day-of-week differences that were statistically significant for a given user and sensor. For example, they might walk significantly more on Friday and Saturday and significantly less on Mondays and Tuesdays. In the original deployment, each day/sensor combination that was significant showed up as a new feed item, which quickly became quite complex to understand and scroll through. Therefore, we rolled up

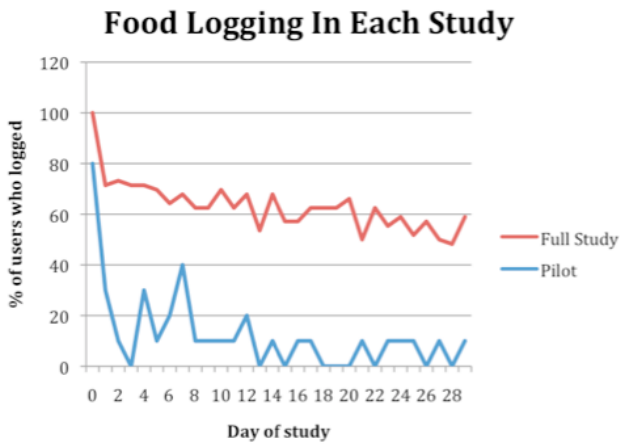

Figure 2: We were able to see a large increase in food logging in our full study after adding silent Android notifications as reminders to log.

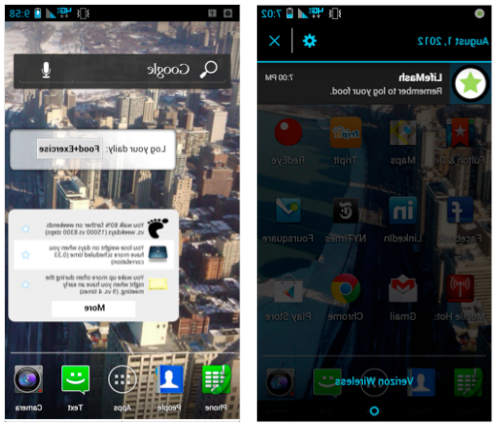

Figure 3: Moving the trigger to log from a static button on the home screen to a notification just fractions of an inch higher in the interface brought about the changes in logging seen in Figure 2.

these items into single observations, stating for example: "You walk more on Fridays and Saturdays, but significantly less on Mondays and Tuesdays."

Using Data to Verify Effects of Design Changes

Once each of these changes was made, we could use the running deployment to validate their effectiveness and continue to tweak the design if necessary.

For the manual logging of Food, we saw a five-fold increase in logging frequency between the pilot and the full study (as shown in Figure 2). Moreover, we saw this high logging frequency continue throughout the duration of the three-month deployment.

We also observed a large increase in the use of the application to view observations. Most usage came directly from the Android notifications when new observations were available and not through manually launching the application, which validated our design to use notifications. Additionally, users commented on the ease of interacting with the system and how they rarely forgot to log since they have a notification to remind them each night.

Finally, we were able to observe an overall increase in use of the system as shown in Figure 4. We hypothesize that this is a virtuous cycle where providing better data to the system provided better and more statistically significant observations, which kept our users coming back to the service over time, thus providing even more data.

Using Data to Create Design Guidelines For Future Services

Beyond just improving our service, the data from our deployments and confirmation of design choices has led to more general guidelines for mobile systems that can inspire the design of future systems.

Most specifically, the findings related to the use of silent notifications demonstrate the power of a very different model of interaction for mobile systems, one in which the system prompts usage. While earlier work using SMS or interruptive reminders showed that users quickly became annoyed [5], our work showed that a silent notification is quickly attended to the next time the user turns on the phone and was not seen as annoying in our interviews after the study. Silent notifications that lead to quick actions such as logging food on a few 7-point scales can increase engagement with a service which then leads to increased use of the service overall.

This is a design pattern that we plan to deploy in future work outside of the health domain.

Discussion

Instrumentation of mobile applications, combined with qualitative data from interviews and voicemail logs, is allowing new methods for design to emerge.

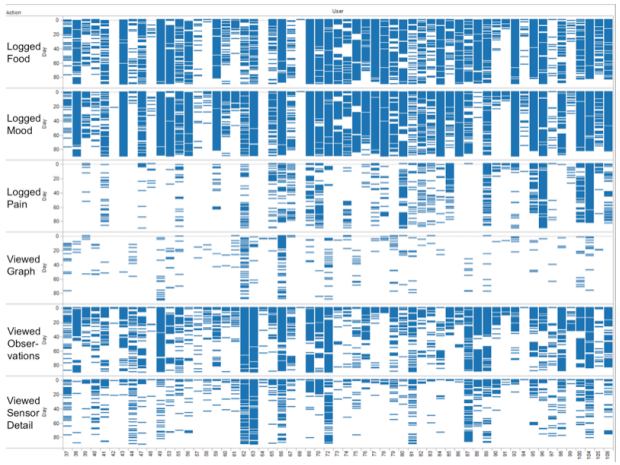

Figure 4: Overall use of each feature of the service over the 90 days of the trial. Each column is a user.

Understanding how a system is being used, from a combination of quantitative and qualitative data, can lead to design insights that can quickly be implemented and validated in real-world deployments to a large number of users. In this project, we were able to be successful not just by blindly analyzing quantitative data and performing A/B tests, but by using the qualitative data from our interviews to deeply understand what was happening that led to the usage patterns that we observed.

This combination of qualitative and quantitative data is critical to deciding on a path forward. We could have easily seen the minimal usage in the first trial as a signal that the concept was not worth pursuing further. However, the qualitative data told a different story of the deep new insights our participants were able to make into their wellbeing and the focused behavior changes that could emerge. By understanding how interaction with our service could be improved, we were able to develop a very successful iteration that easily overcame the lack of use found in the initial pilot system.

We are already discussing how the findings around the use of notifications can transfer into other applications and the methods for instrumentation and analysis that we deployed for this project will be used in future projects to understand how new mobile services are adopted in everyday use.

Acknowledgements

We would like to thank Cristobal Viedma for his work on the implementation of the pilot system and help with the field trial in Stockholm. We would also like to thank Peter Stephenson, and the Georgia Tech IPAT group for their help with the larger field study.

References

[1] Frank R. Bentley and Crysta J. Metcalf. 2007. Sharing motion information with close family and friends. In Proc CHI '07.

[2] Tollmar, K.; Bentley, F.; Viedma, C. 2012. "Mobile Health Mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device," In Proc Pervasive Health 2012

[3] Frank Bentley and Konrad Tollmar. 2013. The Power of Mobile Notifications to Increase Wellbeing Logging Behavior. In Proc CHI '13.

[4] Younghee Jung, Per Persson, and Jan Blom. 2005. DeDe: design and evaluation of a context-enhanced mobile messaging system. In Proc. CHI '05.

[5] Tsai, C., Lee, G., Raab, F., Norman, G., Sohn, T., Griswold, W., Patrick, K. 2007. Usability and Feasibility of PmEB: A Mobile Phone Application for Monitoring Real Time Caloric Balance. Mobile Networks and Applications. 12(2-3), pp 173-184.